Filling Then Spatio-Temporal Fusion for All-Sky MODIS Land Surface Temperature Generation

Yijie Tang, Qunming Wang, and Peter M. Atkinson

***Abstract*—The thermal infrared band of the moderate resolution imaging spectroradiometer (MODIS) onboard the Terra/Aqua satellite can provide daily, 1 km land surface temperature (LST) observations. However, due to the influence of cloud contamination, spatial gaps are common in the LST product, restricting its application greatly at the regional scale. In this article, to deal with the challenge of large gaps (especially complete data loss) in MODIS LST for local monitoring, a filling then spatio-temporal fusion (FSTF) method is proposed, which utilizes another type of product with all-sky coverage, but coarser spatial resolution (i.e., the 7 km China Land Data Assimilation System (CLDAS) LST product). Due to the great temporal heterogeneity of LST, temporally closer auxiliary MODIS LST images are considered to be preferable choices for spatio-temporal fusion of CLDAS and MODIS LST time-series. However, such data are always abandoned inappropriately in conventional spatio-temporal fusion if they contain gaps. Accordingly, pregap filling is performed in FSTF to make fuller use of the valid information in temporally close MODIS LST images with small gaps. Through evaluation in both the spatial and temporal domains for three regions in China, FSTF was found to be more accurate in reconstructing MODIS LST images than the original spatio-temporal fusion methods. FSTF, thus, has great potential for updating the current MODIS LST product at the global scale.**

***Index Terms*—Gap filling, land surface temperature (LST), moderate resolution imaging spectroradiometer (MODIS), spatio-temporal fusion.**

I. Introduction

LAND surface temperature (LST) is an important physical quantity that can be used for studying the interaction between the Earth's surface and the atmosphere [1], [2], [3]. Until now, LST has been applied widely in research on regional drought monitoring [4], [5], land surface evapotranspiration estimation [6], [7], and the heat island effect [8], [9]. Remote sensing provides great potential for large-scale LST monitoring, as LST can be retrieved from the thermal infrared (TIR) band of sensors onboard several satellites (e.g., the Landsat and Terra/Aqua satellites) [10], [11], [12], [13].

Although theories for LST retrieval have been developed and applied widely using various sensors, the LST estimated using satellite sensor data has an obvious defect. Existing studies demonstrate that about 67% of the Earth's surface is contaminated by cloud at any one time [14]. As TIR wavelengths cannot penetrate cloud, some level of data loss in acquired images is to be expected, caused by poor imaging environments. This characteristic restricts greatly the applications of satellite-derived LST products, such as in global climate change studies which require spatial continuity of data. For example, the moderate resolution imaging spectroradiometer (MODIS) sensor onboard Terra, as an important data source for generating global LST, provides a 1 km LST product (i.e., MOD11A1), the scale of which is appropriate for LST research missions at the regional scale. However, the unpredictable imaging environment cannot guarantee data integrity, resulting in spatial information loss in daily MODIS LST. Thus, to provide an all-sky MODIS LST product, the missing information in the daily MODIS LST needs to be reconstructed.

Generally, for small areas of information loss, satisfactory restoration results can be obtained by adopting spatial reconstruction methods in remote sensing [15], [16]. As these methods reconstruct the missing area based only on spatially complete images from the same data source, when the missing area is large the uncertainty in reconstructing the missing information can also be large. Thus, to ensure high accuracy of all-sky MODIS LST generation, other auxiliary data and reconstruction methods are required when the information loss is large (especially completely lost in a local area of interest). Apart from the above-mentioned LST products made using thermal remote sensing, land surface models (LSMs) for land data assimilation, including the China Land Data Assimilation System (CLDAS) [17] and Global Land Data Assimilation System [18], can also provide large area LST data with a relatively coarse spatial resolution. As these datasets are produced by combining observations from a variety of sources, such as ground and satellite sensor observations, these systems can also provide an all-sky spatially complete LST product. Due to this advantage, LSMs have great potential to supplement the missing information in optical image-derived products, and assist the reconstruction of MODIS LST [19]. Thus, in this article when reconstructing MODIS LST with large areas of data loss, a typical LSM is considered, such as the CLDAS data which can provide 7 km spatially seamless hourly LST products covering the Asian area.

Manuscript received 12 October 2022; revised 19 December 2022; accepted 8 January 2023. Date of publication 11 January 2023; date of current version 24 January 2023. This work was supported by the National Natural Science Foundation of China under Grant 42171345, Grant 42222108, Grant 41971297, and Grant 42221002. *(Corresponding author: Qunming Wang.)*

Yijie Tang and Qunming Wang are with the College of Surveying and Geo-Informatics, Tongji University, Shanghai 200092, China (e-mail: tangyijie@sina.com; wqm11111@126.com).

Peter M. Atkinson is with the Faculty of Science and Technology, Lancaster University, Lancaster LA1 4YR, U.K., and also with the Department of Geography and Environment, University of Southampton, Southampton SO17 1BJ, U.K. (e-mail: pma@lancaster.ac.uk).

Digital Object Identifier 10.1109/JSTARS.2023.3235940

Specifically, spatio-temporal fusion can help to be used to reconstruct the missing MODIS LST based on spatially complete MODIS-CLDAS LST image pairs at other times, and a CLDAS LST image at the prediction time.

Over the past decades, various categories of spatio-temporal fusion approaches were developed [20], [21], [22], [23], [24], [25]. Generally, the spatial weighting-based and the spatial unmixing-based methods are the two earliest types of spatio-temporal fusion methods. Specifically, the common practice for spatial weighting-based methods is to predict the center fine pixels by quantifying the contribution of the neighboring spectrally similar pixels using a weighting function. The spatial and temporal adaptive reflectance fusion model (STARFM) [26] is regarded as the classic example of this type. Based on STARFM, the enhanced STARFM (ESTARFM) algorithm [27], the Fit-FC method [28], the spatial temporal adaptive algorithm for mapping reflectance change [29], and the spatial weighting-based virtual image pair-based spatio-temporal fusion (VIPSTF-SW) [30] approaches were further developed.

Spatial unmixing-based methods estimate the value of the fine pixels by solving the reflectance of all object classes through different kinds of unmixing models. This category of method was developed on the basis of the multisensory, multiresolution technique proposed by Zhukov [31]. By applying different constraints and auxiliary data to the unmixing model, various spatial unmixing-based algorithms were proposed [32], [33], [34], [35]. Additionally, recently developed methods include hybrid methods integrating the advantages of the spatial weighting and the spatial unmixing-based methods [36], [37], [38], [39], together with learning-based methods such as the sparse-representation-based spatiotemporal reflectance fusion model (SPSTFM) [40] and the wavelet-artificial intelligence fusion approach [41]. Except for conventional machine learning methods, deep learning-based models were also developed due to their advantages in describing the complex relationship between coarse and fine spatial resolution images. So far, spatio-temporal fusion methods based on different improved versions of convolutional neural networks [42], [43], [44], [45] and generative adversarial networks (GANs) [46], [47], [48] were proposed. Some examples are the multistage remote sensing image spatio-temporal fusion network (MSFusion) [45] and the GAN-based spatiotemporal fusion model (GAN-STFM) [46].

Although spatio-temporal fusion provides practical means for reconstructing large areas of information loss in MODIS LST, their reliability is limited by the availability of auxiliary data. It is acknowledged that due to strong temporal heterogeneity, LST changes greatly over time, and images with closer acquisition times tend to have greater similarity. Thus, in spatio-temporal fusion, fine spatial resolution images with a shorter time interval are considered as a preferable choice for the auxiliary data. Nevertheless, due to frequent cloud cover, there exists great difficulty in finding temporally close images with spatially complete coverage. Although the common practice to search for cloud-free images with a longer time interval can avoid the impact of data quality, the prediction may contain large uncertainties as there can be great LST changes between the prediction and known times. Actually, the impacts of cloud on remote sensing images vary, and temporally close images with a small patch of data missing have the potential to provide significant auxiliary information to the prediction. Thus, for the reconstruction of MODIS LST with large areas of data loss, there is a great need to develop spatio-temporal fusion methods to take full advantage of temporally close LST images with gaps.

To take fuller advantage of the available MODIS LST data, a filling then spatio-temporal fusion (FSTF) method is proposed. Instead of searching for spatially complete, but temporally far MODIS LST as auxiliary data, FSTF considers temporal proximity also, that is, it utilizes the MODIS LST data temporally closest to the prediction time for spatio-temporal fusion. Considering that there probably exist missing spatial data (but with small gaps) in the temporally closest MODIS LST images, gap filling is first applied to keep the integrity of the auxiliary data. Many gap filling methods are available for this task. Amongst the existing filling methods, the most commonly used are hybrid methods that integrate the spatial information of the remaining valid data and the temporal information of auxiliary images acquired at other times (i.e., temporally neighboring image with complete coverage) [49], [50]. Some examples are the neighborhood similar pixel interpolator (NSPI) [51] and the modified NSPI (MNSPI) [52] methods. Additionally, learning-based methods were also developed recently [53], [54], [55], with the unique advantage of characterizing the nonlinear relation between the images with gaps and the auxiliary data.

After gap filling the temporally close MODIS LST data, spatio-temporal fusion is implemented using the gap filled MODIS LST, along with the CLDAS LST at the known and prediction times. FSTF provides a new solution to the reconstruction of MODIS LST with large gaps when there exists difficulty in finding spatially complete auxiliary data and LST changes greatly over time, thus solving the key issue in all-sky MODIS LST generation. Generally, this article makes the following three main contributions.

1) The FSTF method makes fuller use of the valid information in the images acquired temporally close to the prediction time, thus, breaking through the limitation on data integrity for traditional spatio-temporal fusion and enhancing the flexibility of data selection for the case with strong temporal heterogeneity (e.g., the MODIS LST studies in this article).
2) By integrating gap filling and the spatio-temporal fusion techniques effectively, more accurate all-sky MODIS LST can be reconstructed.
3) The generated all-sky LST has the potential to provide hourly MODIS-like LST by inheriting the hourly temporal resolution of CLDAS LST, facilitating fine temporal resolution (e.g., diurnal) LST change monitoring. The hourly, 1 km LST has great potential for various studies based on the need for dynamic LST at the regional scale.

The remainder of this article is organized into four sections. In Section II, the data and the proposed FSTF method are introduced. Section II-A and B present the three study areas and the two categories of data utilized in this research. Section II-C introduces the principles of the gap filling and spatio-temporal fusion methods employed in the experiments. Section II-D, E,

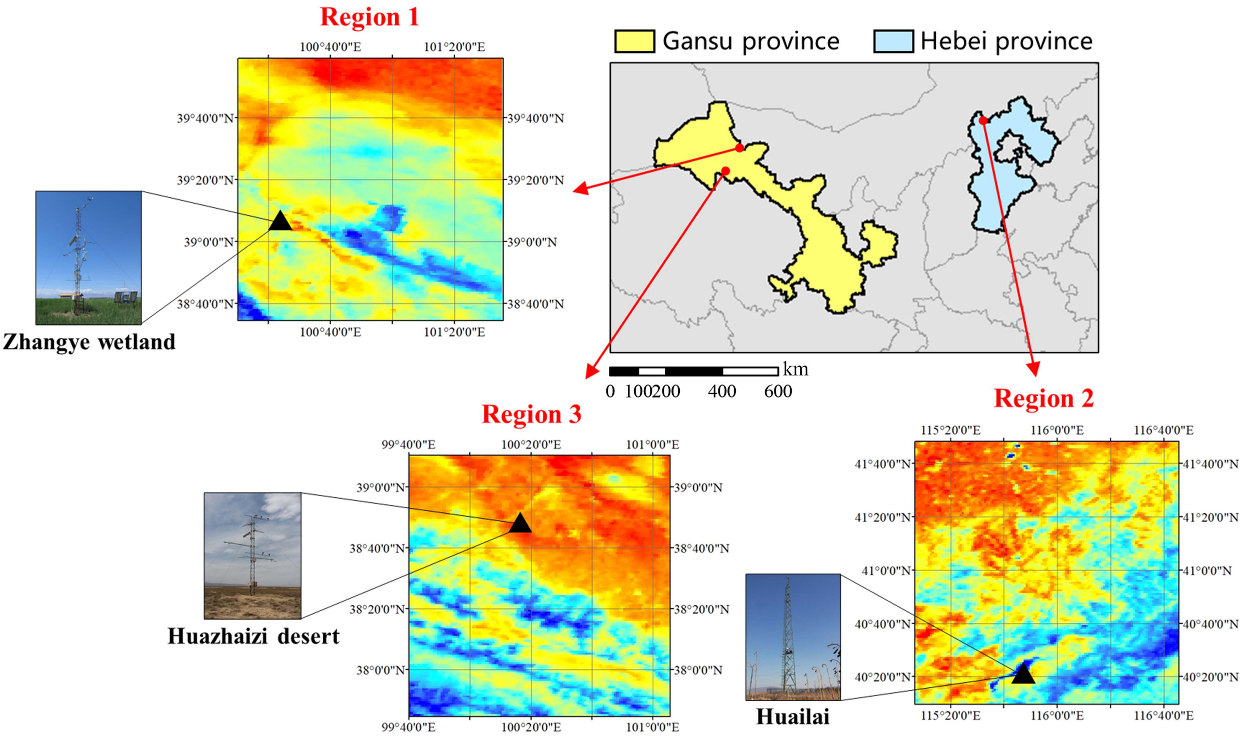


Fig. 1. Three study areas.

and F present the principle, theoretical basis, and implementation of the proposed FSTF method, respectively. In Section III, the effectiveness of FSTF is validated in both the spatial and temporal domains. Section IV further discusses the potential and limitations of FSTF. Section V concludes this article.

II. Methods

A. Study Area

The experiments for this research were implemented in three regions in North China, each covering a spatial extent of 140 km × 140 km. Each area is covered by one meteorological station providing ground-based measurements. The location of the three regions and the corresponding meteorological stations are shown in Fig. 1. Specifically, the three stations are Zhangye wetland station (100°26′47.04″E, 38°58′30.36″N) [56] located in Gansu province (Region 1), Huailai station (115°47′32.28″E, 40°21′26.64″N) [57] located in Hebei province (Region 2) and Huazhaizi desert station (100°19′12.36″E, 38°45′57.24″N) [56] located in Gansu province (Region 3). The automatic meteorological stations are all installed on towers, providing datasets including air temperature, wind speed, precipitation, and four-component radiation every 10 min. For the three towers in Regions 1–3, the underlying surfaces are reed wetland, corn belt, and piedmont desert, respectively.

B. Data

The aim of this research was to reconstruct MODIS LST images, with the assistance of the CLDAS LST product. The datasets in this experiment, therefore, included: the MODIS LST product (MOD11A1), CLDAS LST product, and ground-based LST.

1) MODIS and CLDAS LST data: MODIS LST is provided by the MOD11A1 daily surface temperature product (version 6) (MODIS/Terra LST/emissivity daily L3 global 1 km SIN grid product), which can be obtained from https://search.earthdata.nasa.gov/. This product was generated by applying the split window algorithm to bands 31 and 32 of MODIS onboard Terra. The MOD11A1 product provides 1 km observations, including daytime and nighttime LST, quality indicators, and observation times. When MODIS data are acquired, they are first reprojected to the same coordinate system as that of CLDAS (WGS84) by the MODIS Reprojection Tool (MRT). Then, to obtain effective MODIS data, the acquired MODIS LST is filtered according to the 8-bit byte quality control (QC) flag of pixels in the QC layer. Specifically, pixels with the flag "cloud" and "average LST error > 3 K" are considered to be invalid data. Thus, the MODIS LST pixels with fine data quality can be selected.

CLDAS is a land surface data assimilation system, which can provide a large number of land surface observation products, such as air temperature, air pressure, soil moisture, and LST. Different from MODIS, the CLDAS product integrates ground observations provided by automatic meteorological stations, numerical analysis/forecast products provided by the European Centre for Medium-Range Weather Forecasting, and many other products. The LST product of CLDAS can provide 7 km × 7 km observations covering the Asian area. In terms of temporal resolution, CLDAS provides hourly LST observations, which have the potential to match accurately the acquisition times of MODIS LST images. By examining the acquisition times of the MODIS LST time-series in the study period, CLDAS

LST at universal time coordinated (UTC) 4 a.m., 3 a.m., and 4 a.m. were selected for Regions 1–3, respectively. Also, the fused LST was evaluated using the ground-based LST at the corresponding time. The CLDAS LST can be obtained from http://data.cma.cn/data.

2) Ground-based measurements: Considering that no real reference exists for the reconstructed MODIS LST, ground-based measurements (i.e., in situ data) were used for evaluating the accuracy of the LST time-series data, which is a common strategy [58], [59]. The ground measurements were acquired from the automatic meteorological station data provided by the Institute of Tibetan Plateau Research (http://data.tpdc.ac.cn). The three meteorological stations collect four-component radiation every 10 min, providing important variables for the acquisition of ground-based LST. Generally, the ground-based LST can be calculated according to the Stefan–Boltzmann law based on the surface upwelling and atmospheric downwelling longwave radiation

$$T_{\mathrm{s}} = \left(\frac{L^{\uparrow} - (1-\varepsilon_b)L^{\downarrow}}{\sigma\varepsilon_b}\right)^{1/4} \tag{1}$$

where T_s is the derived ground-based LST, and σ is the Stefan–Boltzmann's constant ($5.67\times10^{-8}\mathrm{Wm^{-2}K^{-4}}$). $L^{\uparrow}$ and $L^{\downarrow}$ are surface upwelling and atmospheric downwelling longwave radiation, respectively, and ε_b is the broadband emissivity, which can be estimated by

$$\varepsilon_b = 0.2122\cdot\varepsilon_{29} + 0.3859\cdot\varepsilon_{31} + 0.4029\cdot\varepsilon_{32} \tag{2}$$

where ε_{29}, ε_{31} and ε_{32} are narrowband emissivities of MODIS bands 29, 31, and 32, which can be obtained from the MODIS/Aqua LST/3-band emissivity daily L3 global 1 km product (MYD21A1) [60], [61].

For the three study areas, the reconstructed MODIS LST was evaluated using the in situ LST at the same acquisition time.

C. Gap Filling and Spatio-Temporal Fusion Methods

FSTF implements gap filling first and then applies spatio-temporal fusion using CLDAS LST. This section introduces the gap filling and spatio-temporal fusion methods used in this research.

1) Gap filling (MNSPI): Gap filling aims to reconstruct the spatial information loss caused by cloud or other factors. Until now, methods utilizing both spatial and temporal correlations together are considered to be more reliable compared with methods using either of the two. Typically, for the MNSPI approach, a spatial-based prediction is first estimated according to spatially neighboring information of the data with gaps, and then a temporal-based prediction is made with the assistance of the information in the auxiliary image. For final prediction, the spatial-based and temporal-based predictions are weighted according to the spatial and temporal heterogeneity.

2) Spatio-temporal fusion (STARFM, ESTARFM, VIPSTF-SW, and SPSTFM): Spatio-temporal fusion aims at obtaining fine spatial and temporal resolution images by combining the advantages of images with different resolutions. Generally, the basic mechanisms for predicting the fine spatial resolution image $\mathbf{F}_p$ by spatio-temporal fusion can be summarized using the following framework [30], [37]:

$$\begin{aligned}\hat{\mathbf{F}}_p &= \mathbf{F}_k + \Delta\mathbf{F}_{k\to p}\\ &= \mathbf{F}_k + f(\Delta\mathbf{C}_{k\to p})\end{aligned} \tag{3}$$

where the prediction of $\mathbf{F}_p$ includes two terms: the known fine image $\mathbf{F}_k$ and the fine increment $\Delta\mathbf{F}_{k\to p}$ (i.e., temporal change) to be estimated. The first term makes use of the available fine spatial resolution information directly, while the second term predicts fine spatial resolution change information from the available coarse spatial resolution data. The estimation of $\Delta\mathbf{F}_{k\to p}$ is the core of spatio-temporal fusion, which is calculated by applying different downscaling operators f to the coarse spatial resolution increment $\Delta\mathbf{C}_{k\to p}$. For STARFM, f is a weighting function considering together the spatial, temporal, and spectral differences of neighboring pixels [26]. Based on STARFM, a conversion coefficient is used in ESTARFM to describe this relationship more explicitly [27]. For VIPSTF-SW, spatial weighting-based fusion is performed using a virtual image pair that is generated by applying a linear transformation to the original image pair [30]. For SPSTFM, $\Delta\mathbf{F}_{k\to p}$ is estimated by training a dictionary-pair of patches between two coarse and fine difference image pairs [40].

D. Proposed FSTF Method

In conventional spatio-temporal fusion, when there exist spatial gaps in temporally adjacent MODIS pixels, the image is completely discarded and cloud-free auxiliary images at a more distant time are further adopted. However, considering the great heterogeneity of LST in the temporal dimension, this practice can amplify the uncertainty of spatio-temporal fusion as greater LST changes may occur when there is a large time interval between the known and prediction times. Thus, it is necessary to make fuller use of the important effective information in temporally adjacent images, even when they are contaminated by cloud. Actually, the area of influence of cloud cover in remote sensing images varies greatly. For images with small cloud coverage, abandoning the whole image leads to a significant waste of valuable auxiliary data. Thus, to address this problem, the FSTF algorithm is developed here. Note that this method aims to reconstruct MODIS LST images with a large area of information loss (especially for completely missing data). The process of FSTF (green line) is shown in Fig. 2, presented with a comparison to traditional spatio-temporal fusion (red line).

For FSTF, to predict the MODIS LST at t_p, auxiliary image pairs on other dates need to be selected. Suppose that there are four image pairs acquired before and after the acquisition time of t_p. More precisely, t_m and t_n are temporally close to t_p, while the MODIS LST images contain data loss to some extent. Image pairs acquired at t_k and t_l, however, are temporally further from the predicted MODIS LST, with complete spatial coverage. For traditional spatio-temporal fusion, image pairs acquired at t_k and t_l are applied directly. Due to the great LST change occurring from the known to prediction times, however, the reliability of prediction remains to be examined. Alternatively, for FSTF, MODIS LST images acquired at t_m and t_n are first reconstructed using gap filling methods. Then, the complete MODIS and CLDAS LST image pairs are included in the spatio-temporal

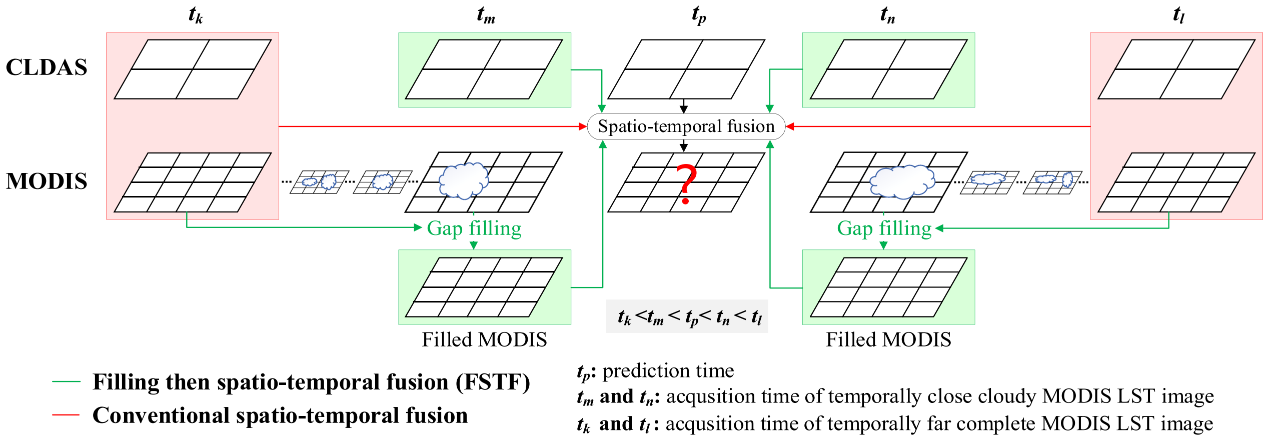

t_k
t_m
t_p
t_n
t_l
CLDAS
MODIS
Spatio-temporal fusion
Gap filling
Gap filling
Filled MODIS
Filled MODIS
$t_k < t_m < t_p < t_n < t_l$
Filling then spatio-temporal fusion (FSTF)
Conventional spatio-temporal fusion
t_p: prediction time
t_m and t_n: acqusition time of temporally close cloudy MODIS LST image
t_k and t_l: acqusition time of temporally far complete MODIS LST image

Fig. 2. Processing involved in the proposed FSTF and conventional spatio-temporal fusion methods.

fusion to make a final prediction. It is noted that the proposed FSTF is a class of method that can be implemented by applying different gap filling and spatio-temporal methods.

When reconstructing the MODIS LST time-series, we divided the acquired MODIS LST images into three classes according to the missing area: no gaps, small gaps (missing area less than 40%), and large gaps (missing area more than 40%). For the small gap case, a gap filling method (i.e., MNSPI) was applied to reconstruct the missing information with the temporally closest, spatially complete MODIS LST images. For the large gaps case, spatio-temporal fusion was implemented based on the temporally closest MODIS LST image, which refers to either the observed MODIS LST image with no gaps or filled data for the small gaps case. By applying the above-mentioned process, all-sky MODIS LST time-series images can be reconstructed.

E. Comparison Between FSTF and Traditional Spatio-Temporal Fusion

For further examination of the rationale of FSTF, a theoretical comparison between the FSTF and traditional spatio-temporal fusion is presented. Suppose that there is a MODIS LST image $\mathbf{M}_m$ acquired at a time close to the MODIS LST $\mathbf{M}_p$ to be predicted, but with a small area of spatial information loss. Also, the temporally further, spatially complete MODIS LST $\mathbf{M}_k$ is available. Traditionally, spatio-temporal fusion requires spatially complete auxiliary data. Accordingly, the prediction for $\mathbf{M}_p$ is

$$\hat{\mathbf{M}}_p = \mathbf{M}_k + f_1(\Delta\mathbf{C}_{k\to p}) \tag{4}$$

where f_1 is the downscaling function in spatio-temporal fusion, and $\Delta\mathbf{C}_{k\to p}$ is the CLDAS LST increment from t_k to t_p.

FSTF conducts spatio-temporal fusion based on the temporally close MODIS LST image $\mathbf{M}_m$, even if there exists data loss. Thus, the prediction for $\mathbf{M}_p$ is

$$\begin{aligned}\hat{\mathbf{M}}'_p &= \hat{\mathbf{M}}_m + f_1(\Delta\mathbf{C}_{m\to p})\\ &= \mathbf{M}_{m\_valid} + \hat{\mathbf{M}}_{m\_missing} + f_1(\Delta\mathbf{C}_{m\to p})\end{aligned} \tag{5}$$

where $\mathbf{M}_{m\_valid}$ denotes the data for the valid area of $\mathbf{M}_m$ and $\hat{\mathbf{M}}_{m\_mis\sin g}$ is the missing area of $\mathbf{M}_m$ required to be estimated. It is noted that there is no overlap between $\mathbf{M}_{m\_valid}$ and $\hat{\mathbf{M}}_{m\_mis\sin g}$. By applying gap filling based on $\mathbf{M}_k$, (5) can be updated as follows:

$$\begin{aligned}\hat{\mathbf{M}}'_p = {} & \mathbf{M}_{m\_valid} + \mathbf{M}_{k\_mis\sin g} + f_2(\Delta\mathbf{M}_{k\to m\_valid})\\ & + f_1(\Delta\mathbf{C}_{m\to p}).\end{aligned} \tag{6}$$

In (6), $\mathbf{M}_{k\_mis\sin g}$ are the valid data in $\mathbf{M}_k$ that share the same geographical location with the missing area in $\mathbf{M}_m$, and $\Delta\mathbf{M}_{k\to m\_valid}$ is the MODIS LST increment from t_k to t_m for the valid area. f_2 is a spatial interpolation algorithm.

For comparison between traditional spatio-temporal fusion and FSTF, the traditional version in (4) is altered to

$$\begin{aligned}\hat{\mathbf{M}}_p &= \mathbf{M}_k + f_1(\Delta\mathbf{C}_{k\to m}) + f_1(\Delta\mathbf{C}_{m\to p})\\ &= \mathbf{M}_{k\_mis\sin g} + \mathbf{M}_{k\_valid} + f_1(\Delta\mathbf{C}_{k\to m}) + f_1(\Delta\mathbf{C}_{m\to p})\end{aligned} \tag{7}$$

where $\Delta\mathbf{C}_{k\to m}$ and $\Delta\mathbf{C}_{m\to p}$ are the CLDAS LST increments from t_k to t_m and t_m to t_p, respectively. It is noted that f_1 here should be a linear function, which is in accordance with the four spatio-temporal fusion methods applied in this research. Through comparison between (7) with (6), it is found that there are two terms differing, that is, $\mathbf{M}_{k\_valid} + f_1(\Delta\mathbf{C}_{k\to m})$ in (7) and $\mathbf{M}_{m\_valid} + f_2(\Delta\mathbf{M}_{k\to m\_valid})$ in (6). For the two constant parts $\mathbf{M}_{k\_valid}$ and $\mathbf{M}_{m\_valid}$, they cover the same spatial area. However, it is clear that compared with $\mathbf{M}_{k\_valid}$, $\mathbf{M}_{m\_valid}$ is temporally closer to the prediction time and, thus, can provide more reliable auxiliary information. The core is to compare $f_1(\Delta\mathbf{C}_{k\to m})$ with $f_2(\Delta\mathbf{M}_{k\to m\_valid})$. First, from the perspective of spatial scale, $f_1(\Delta\mathbf{C}_{k\to m})$ involves a downscaling process with great uncertainty. However, $f_2(\Delta\mathbf{M}_{k\to m\_valid})$ in FSTF is a spatial interpolation algorithm performed at the same fine spatial resolution with the original data, which tends to involve less uncertainty. Second, from the amount of data for prediction, $f_2(\Delta\mathbf{M}_{k\to m\_valid})$ in FSTF needs only to predict the data for pixels in the missing area, but $f_1(\Delta\mathbf{C}_{k\to m})$ needs to predict the data for all pixels in the entire region. It is widely acknowledged that uncertainty normally exists in any prediction process. In summary, we can conclude that FSTF tends to involve less certainty than traditional spatio-temporal fusion.

TABLE I
ACQUISITION TIMES OF MODIS AND CLDAS DATA UTILIZED IN THE EXPERIMENTS

	Case 1	Case 2	Case 3
Experiment 1	22-Mar-2018	9-Jul-2018	1-May-2020
	27-Apr-2018	23-Jul-2018	16-May-2020
	2-May-2018	31-Jul-2018	17-May-2020
	3-May-2018	1-Aug-2018	22-May-2020
	6-May-2018	22-Aug-2018	28-Jun-2020
	Region 1	Region 2	Region 3
Experiment 2	1-Mar-2018- 31-Aug-2018	1-Mar-2018- 30-Jun-2018	1-May-2020- 31-Aug-2020

F. Implementation of the FSTF Method

The specific implementation of FSTF is as follows:

Step 1: The obtained MODIS LST time-series was classified into three categories: complete LST image, LST image with small gaps, and LST image with large gaps.

Step 2: The MNSPI method was applied to reconstruct the information in the MODIS LST images with small gaps based on the temporally closest, complete LST image.

Step 3: For MODIS LST images with large gaps, there are two parts. First, the MNSPI method was applied to reconstruct the missing information in the two MODIS LST images acquired temporally closest, before and after the prediction time. Second, based on the reconstructed MODIS LST images, spatio-temporal fusion was applied to reconstruct the missing MODIS LST image, with the assistance of the corresponding CLDAS LST images at the three times (the prediction time and the two known times of reconstructed MODIS LST data).

III. EXPERIMENTS

The experiments for this research are divided into two parts. In the first experiment in Section III-A, the proposed FSTF method was examined in the spatial dimension. That is, the reconstructed MODIS LST image was evaluated (predicting one LST image at each time) by comparison with the reference image with complete spatial coverage. According to the process of FSTF, Section III-A presents the results of gap filling and spatio-temporal fusion, followed by a comparison between different spatio-temporal fusion methods to provide the FSTF version with the greatest performance for Section III-B. In the second experiment in Section III-B, the performance of FSTF was tested in the temporal dimension. That is, the reconstructed MODIS LST time-series was evaluated based on the temporal profile of each pixel, by referring to the in situ LST measurements. Table I lists the acquisition times of the MODIS and CLDAS data utilized in Experiments 1 and 2.

A. Experiment 1: Evaluation in the Spatial Dimension

To examine the feasibility of FSTF, a comparison experiment was conducted between spatio-temporal fusion (simplified to STF in the experiments) and FSTF. The research areas included in this experiment were selected amongst the three regions introduced in Section III-A. The data are shown in Fig. 3. For Cases 1 and 2, experimental data were selected from Region 1. While for Case 3, data were selected from Region 3. To fully validate the performance of FSTF, both simulated and real missing data were considered in this experiment.

For Case 1, the MODIS LST image on 2 May 2018 was predicted. For STF, MODIS and CLDAS LST image pairs acquired on 22 March 2018 and 6 May 2018 were used, which are 41 and 4 days away from the prediction date, respectively. For FSTF, we simulated gaps for MODIS LST images on 27 April 2018 and 3 May 2018, which are 5 and 1 days away from the prediction date, respectively. Then, MNSPI was implemented to reconstruct the simulated missing data in MODIS LST. Finally, spatio-temporal fusion was conducted to predict MODIS LST on 2 May 2018 based on spatial reconstructed image pairs on 27 April 2018 and 3 May 2018. The available spatially complete MODIS LST image on 2 May 2018 was used as reference for evaluation. Cases 2 and 3 were implemented based on real MODIS and CLDAS data. For Case 2, MODIS LST on 31 July 2018 was predicted. The temporally closest complete MODIS LST images, which were acquired 29 days earlier and 22 days later than the prediction date were included in STF. For FSTF, the temporally closest MODIS LST images with small gaps were used, which were acquired 8 and 1 days away from the prediction dates. For Case 3, to predict the MODIS LST image on 17 May 2020, the spatially complete MODIS LST acquired 16 and 42 days away were used for STF, while MODIS LST data with small gaps acquired 1 and 5 days away were applied in FSTF. Similarly, the available MODIS LST images (with no gaps) on 31 July 2018 and 17 May 2020 were used as references for evaluation in Cases 2 and 3, respectively. For the three cases, the STARFM, ESTARFM, SPSTFM, and VIPSTF-SW methods were applied to STF, while the four corresponding FSTF versions were also tested. It is noted that FSTF is a class of methods composed of a gap filling method and a spatio-temporal fusion method. In this research, FSTF was specified by integrating the MNSPI method and four different spatio-temporal fusion methods (i.e., STARFM, ESTARFM, SPSTFM, and VIPSTF-SW), which are named as STARFM-based FSTF, ESTARFM-based FSTF, SPSTFM-based FSTF, and VIPSTF-SW-based FSTF, respectively.

1) Gap filling: Gap filling is the first step for FSTF, the performance of which can affect the final prediction. The gap filling results for three cases are shown in Fig. 4. Amongst the three cases, quantitative evaluation can be conducted for the simulated experiment (i.e., Case 1), as the spatially complete MODIS LST images at the corresponding time are available. For Case 1, the correlation coefficients (CCs) for the reconstructed MODIS LST images on 27 April 2018 and 3 May 2018 are 0.929 and 0.820, respectively. Moreover, the root mean square errors (RMSEs) are 1.221 K and 1.142 K, respectively. As the performance of MNSPI depends on the size and position of missing area greatly, the accuracy varies in different gap filling cases. Generally, the results of MNSPI for the three cases have a satisfactory performance considering the visual continuity and relatively harmonious hue.

2) Spatio-temporal fusion: The results of FSTF and STF are presented in Fig. 5. Checking the fusion results, the predictions of FSTF are visually more similar to the reference than those for STF in most cases. Specifically, for Case 1, the values of the STF

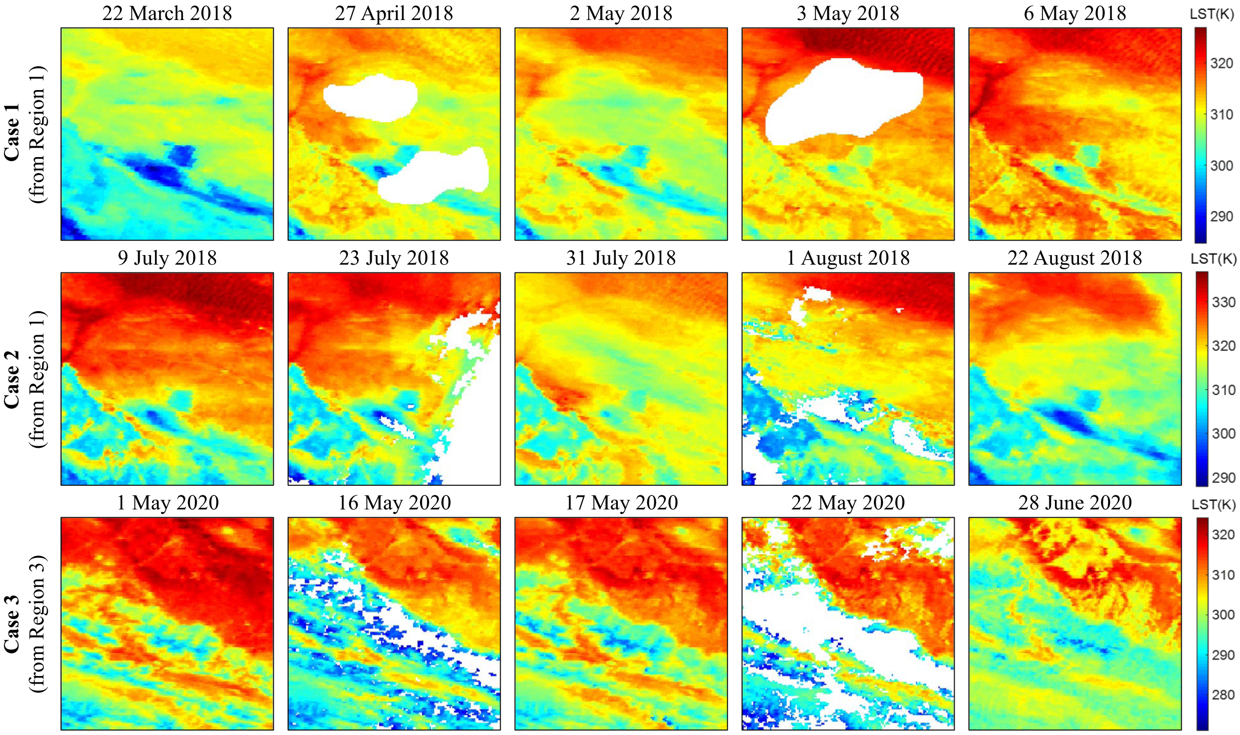


Fig. 3. MODIS LST data used in Experiment 1.

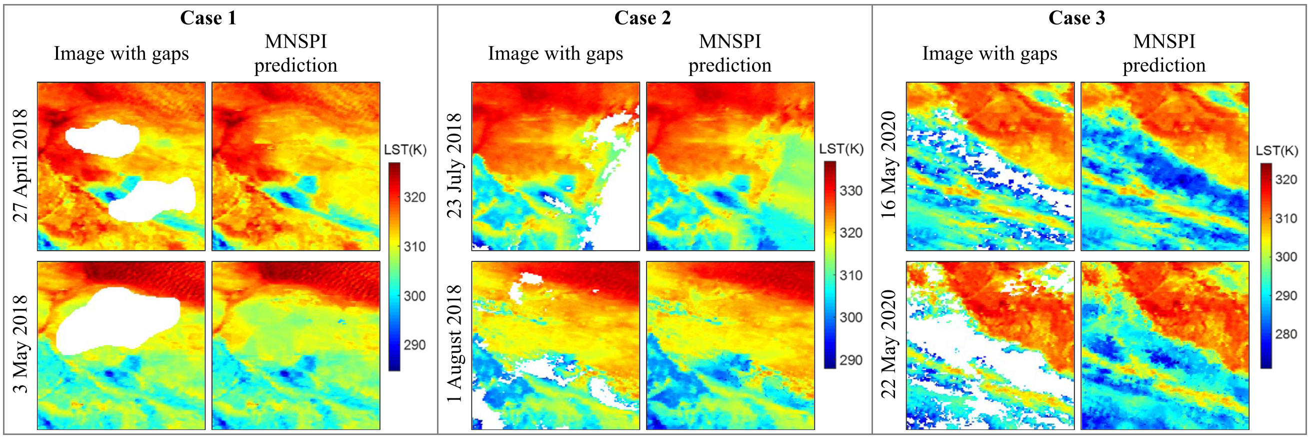


Fig. 4. Gap filling results of the temporally close MODIS LST images with gaps in the three cases in Experiment 1.

predictions tend to be larger than those in the reference, which incorrectly visually present as red, especially in the middle and upper left of the image. Generally, the ESTARFM-based and VIPSTF-SW-based results have a color similar to the reference. For Case 2, almost all the predictions present an overprediction of LST. Compared with the results of STF, the predictions of all three FSTF versions are visually more accurate and they present more similar colors to the reference. Amongst all four versions, the prediction of the VIPSTF-SW-based methods appears to be the closest to the reference visually, and other predictions overestimate the range of high-temperature area. For Case 3, the prediction using FSTF is challenged, as the missing area overlaps greatly in the two auxiliary MODIS LST images. Moreover, although the MODIS LST image pairs involved in STF are temporally further from the prediction date, the MODIS LST image acquired on 1 May 2020 is quite similar to the predicted MODIS LST image due to the irregular variation of LST. Thus, in this case, FSTF may fail to produce more satisfactory results than STF. From visual inspection, however, FSTF presents more satisfactory results for all four versions in the prediction of the upper half of the image, but fails to predict the correct color in the middle of image corresponding to the missing area in the auxiliary images.

Quantitative evaluation for spatio-temporal fusion is implemented based on RMSE and CC, as exhibited in Table II. For Cases 1 and 2, the accuracies of FSTF are obviously greater than for STF. For Case 1, the RMSEs of STARFM-based, ESTARFM-based, SPSTFM-based, and VIPSTF-SW-based

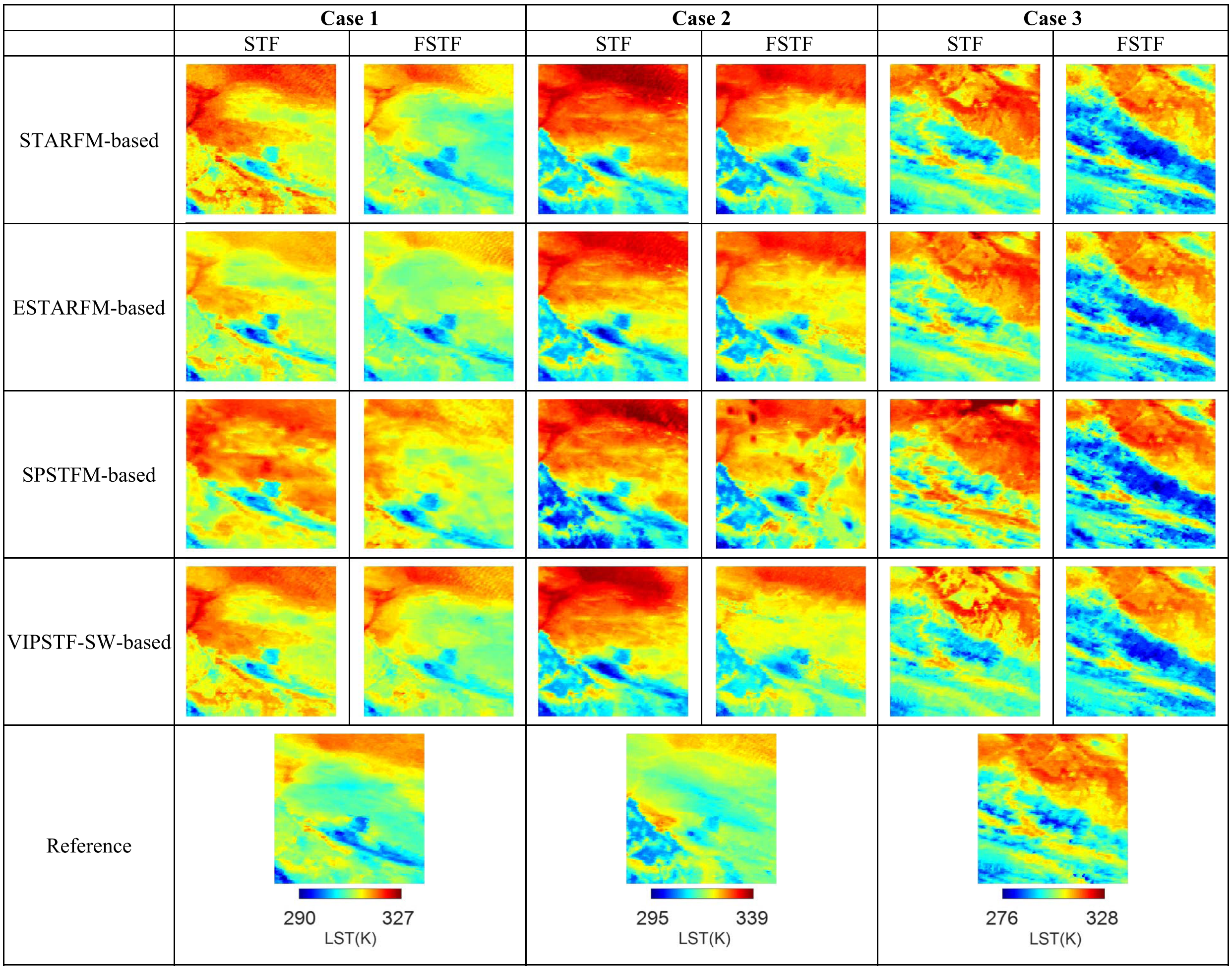


Fig. 5. Results of STF and FSTF in Experiment 1.

TABLE II
ACCURACIES FOR THE THREE CASES IN EXPERIMENT 1

		Case 1		Case 2		Case 3	
		STF	FSTF	STF	FSTF	STF	FSTF
RMSE (K)	STARFM-based	5.734	2.588	9.705	6.725	2.740	5.788
	ESTARFM-based	3.429	2.480	7.967	6.346	**2.601**	5.226
	SPSTFM-based	5.495	**2.071**	8.906	6.255	4.802	7.147
	VIPSTF-SW-based	5.034	2.980	8.783	**4.890**	3.170	5.342
CC	STARFM-based	0.820	0.832	0.674	0.744	0.952	0.914
	ESTARFM-based	0.787	0.832	0.711	0.783	**0.965**	0.921
	SPSTFM-based	0.701	**0.872**	0.691	0.734	0.932	0.887
	VIPSTF-SW-based	0.836	0.868	0.683	**0.784**	0.926	0.914

FSTF predictions are 3.146, 0.949, 3.424, and 2.054 K smaller than for the original methods. In addition, the corresponding CCs of the FSTF versions are 0.012, 0.045, 0.171, and 0.032 larger than for the four original methods. Amongst all the predictions, SPSTFM-based FSTF produces the smallest RMSE of 2.071 K and the largest CC of 0.872. For Case 2, the RMSEs of STARFM-based, ESTARFM-based, SPSTFM-based, and VIPSTF-SW-based FSTF are 2.980, 1.621, 2.651, and 3.893 K smaller than for the corresponding original methods. Overall, VIPSTF-SW-based FSTF produces the smallest RMSE of 4.890 K and largest CC of 0.784. For Case 3, FSTF produces less accurate predictions than the original methods, which corresponds to the visual inspection. Actually, in most cases, LST acquired within a few days tends to be more similar, while LST acquired more than half a month differs more according to the natural variation in temperature. Objectively, when the MODIS LST images were acquired temporally further away, but are more similar in value to the prediction date, FSTF may fail to produce more accurate prediction. This kind of extreme situation, however, is rare in practice, which requires subjective inspection in the auxiliary data selection process. Thus, in spatio-temporal fusion, temporally closer image pairs are still a preferable choice for fusion. To avoid the contingency of the experiment, examination of the temporal dimension will be conducted in Section III-B, with the mission of reconstructing MODIS LST time-series.

3) Comparison between different spatio-temporal fusion methods: In this research, four spatio-temporal fusion methods (i.e., STARFM, ESTARFM, SPSTFM, and VIPSTF-SW) were applied to both STF and FSTF. From visual inspection, the results of the four methods seem to be similar, for both STF and FSTF versions. Checking the quantitative evaluation results, it is

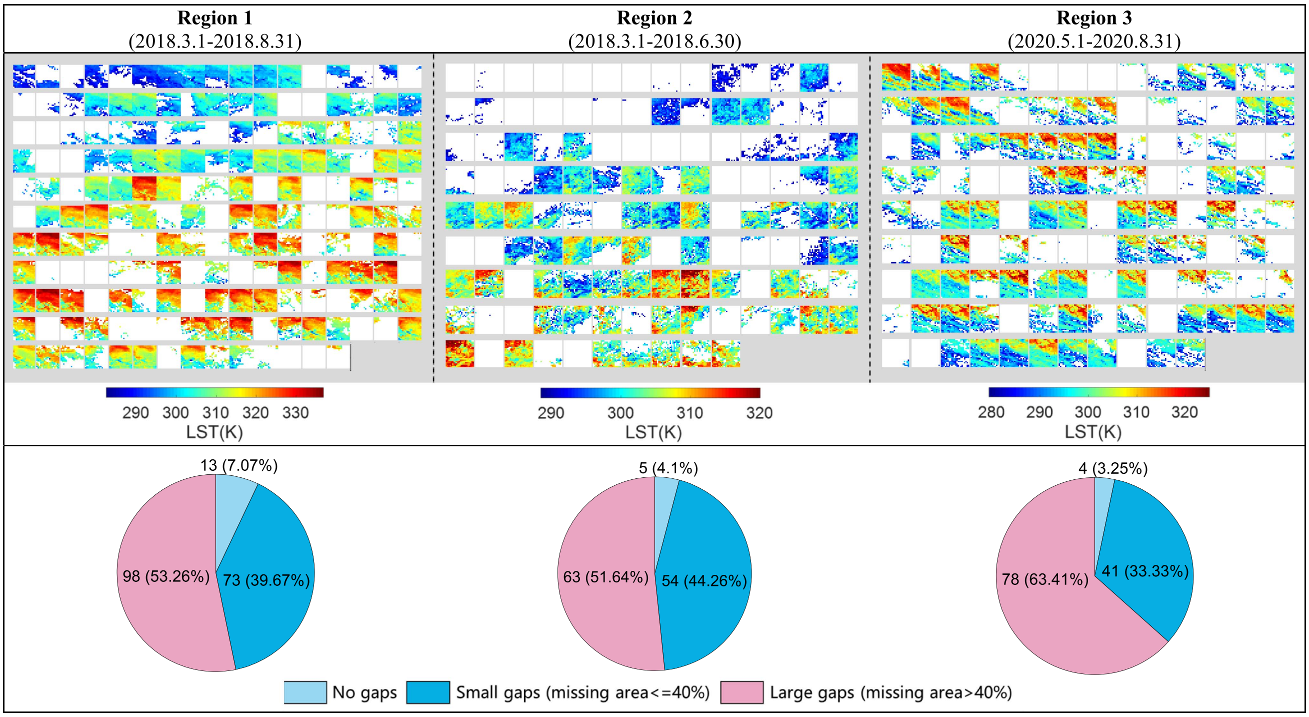


Fig. 6. MODIS LST time-series for the three regions in Experiment 2.

found that the ESTARFM-based and VIPSTF-SW-based methods tend to be more accurate for both the STF and FSTF versions. It is noted that although the predictions of SPSTFM-based FSTF have the largest CC and smallest RMSE for Case 1, its accuracy declines greatly for Cases 2 and 3. While VIPSTF-SW-based FSTF produces the second largest CC of 0.868 for Case 1, and the largest CC and the smallest RMSE for Case 2. Specifically, the CC of VIPSTF-SW-based FSTF for Case 2 is 0.040, 0.001, and 0.050 larger than for STARFM-based, ESTARFM-based, and SPSTFM-based FSTF, respectively. For RMSE, VIPSTF-SW-based FSTF is 1.835 K, 1.456 K, and 1.365 K smaller than for STARFM-based, ESTARFM-based, and SPSTFM-based FSTF, respectively. For Case 3, ESTARFM-based STF produces the largest CC of 0.965 and the smallest RMSE of 2.601 K, while the performance of the four methods in FSTF is relatively similar. Considering the relative performances of the four methods in this examination of the spatial dimension, the VIPSTF-SW-based method will be applied to the following experiments in the temporal dimension.

B. Experiment 2: Evaluation in the Temporal Dimension

To examine the performance of FSTF in the temporal dimension, MODIS LST images covering a period of a few months were selected, as shown in Fig. 6. Specifically, images from 1 March 2018 to 31 August 2018, 1 March 2018 to 30 June 2018 and 1 May 2020 to 31 August 2020 were considered for Regions 1–3, respectively. Checking the LST image types defined in Section II-D for the three regions, there are 13, 5, and 4 images with no gaps in the three regions, with a proportion of 7.07%, 4.1%, and 3.25% amongst all available MODIS LST images for Regions 1–3, respectively. Obviously, as cloud contamination tends to be a normal phenomenon, the number of spatially complete MODIS LST is limited. Amongst all three regions, the number of images with large gaps tends to be the largest, occupying 53.26%, 51.64%, and 63.41% for Regions 1–3, respectively. Considering the composition of the three types of images for the three regions, reconstruction of the MODIS LST time-series in Region 3 is the most challenging. Partial reconstruction results of the VIPSTF-SW-based STF and FSTF methods are shown in Fig. 7. Generally, the reconstruction results have a fine spatial continuity and accord with the law of the natural change of LST. For several days, the reconstruction results of STF and FSTF differ greatly, such as the results on 20 August for Region 1, 4 April for Region 2, and 21 July for Region 3.

To quantify the reconstruction accuracy for the three regions, the in situ LST measurements for Zhangye wetland station, Huailai station, and Huazhaizi desert station were applied for Regions 1–3, respectively, as shown in Fig. 8. The left column presents the predictions for STF and FSTF at the location of the stations and the in situ measurements. To present the differences between the predictions of STF and FSTF more clearly, the absolute difference between the prediction and the in situ measurements is shown in the right column (Fig. 8). Generally, the prediction of FSTF is closer to the measured LST than that for STF for the three regions. For Region 1, as the number of spatially complete MODIS LST images is relatively large, the predictions of STF and FSTF appear to be similar on most dates. As can be seen from Fig. 8(b), however, the absolute difference for FSTF on the first half of the dates is smaller than that for STF, and the performances of these two methods in the second

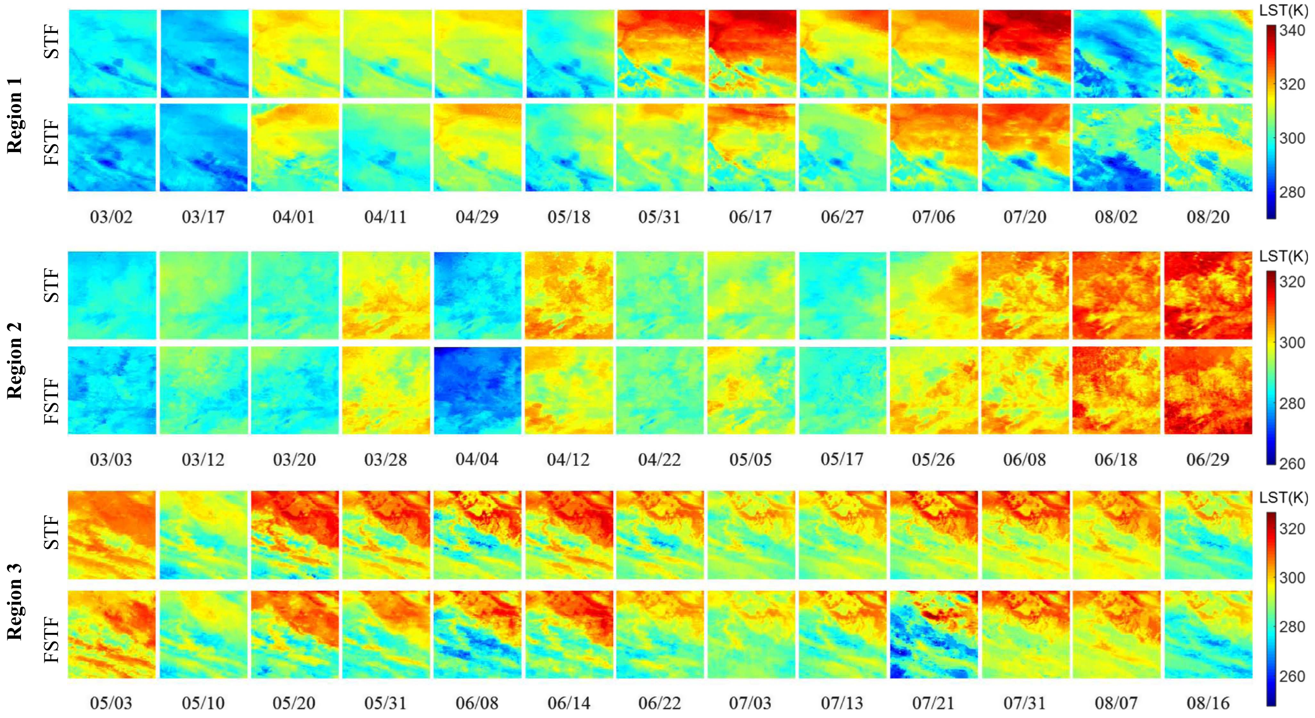
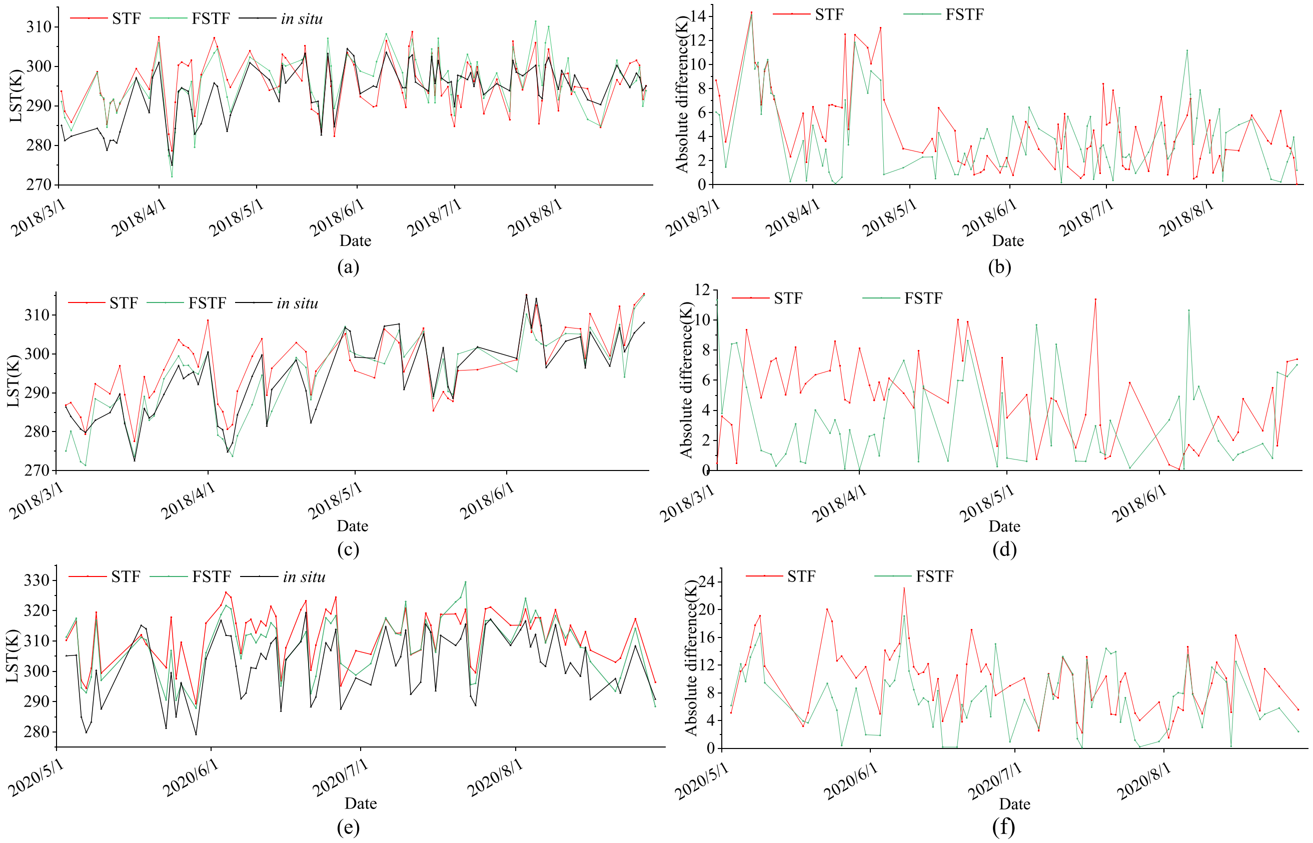


Fig. 7. MODIS LST time-series reconstruction results (partial) in Experiment 2.



(a)

(b)

(c)

(d)

(e)

(f)

Fig. 8. Accuracy for reconstruction of daily MODIS LST. (a) Predicted LST for Region 1. (b) Absolute difference for Region 1. (c) Predicted LST for Region 2. (d) Absolute difference for Region 2. (e) Predicted LST for Region 3. (f) Absolute difference for Region 2.

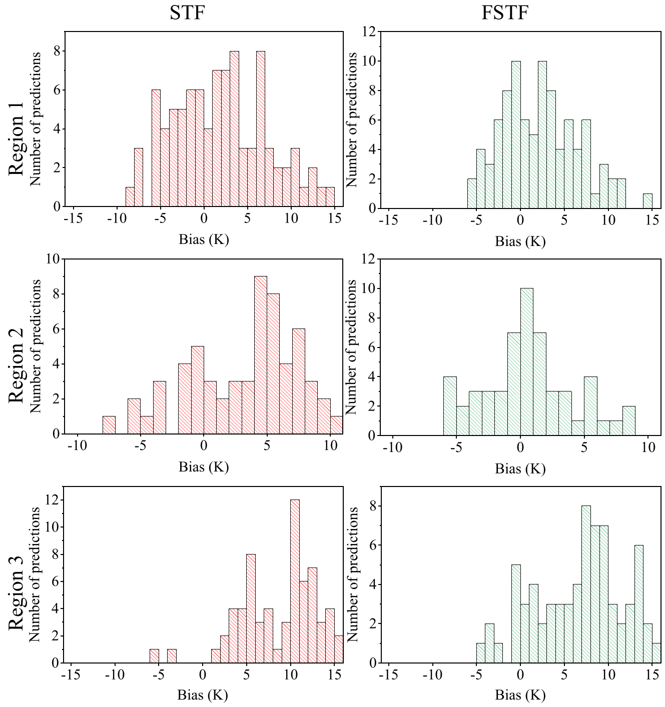


Fig. 9. Error distributions for the reconstruction of daily MODIS LST.

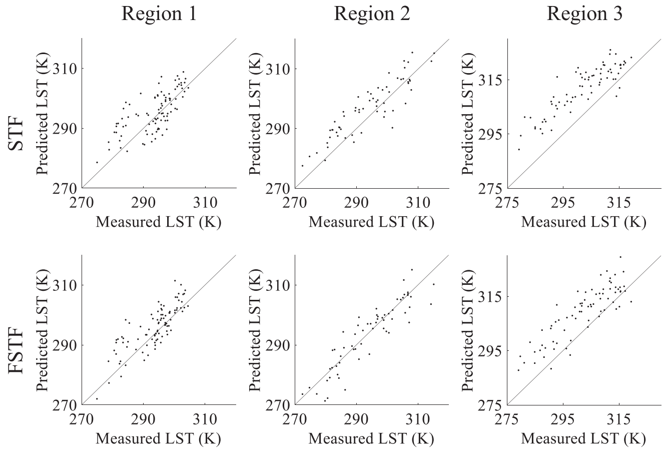


Fig. 10. Scatter plots for the reconstruction of daily MODIS LST.

half of the dates are close to each other. For Region 2, it can be noted in Fig. 8(c) that the green line is closer to the in situ LST measurements, indicating the greater performance of FSTF. Again, the prediction of FSTF also presents a smaller absolute difference in Fig. 8(d). For Region 3, both Fig. 8(e) and (f) indicate that the predicted LST of FSTF is closer to the measured LST on most dates.

To examine the overall performance, the error distributions and scatterplots between the predictions and in situ data are shown in Figs. 9 and 10, respectively. In Fig. 9, it is noted that for all three regions, the error of FSTF is closer to zero compared with STF. In Fig. 10, for all three regions, the scatter plots of FSTF against the in situ data are closer to the $y = x$ line and are more aggregated compared with those for STF and the in situ data. To further evaluate the overall accuracy, the mean absolute error (MAE), RMSE, and coefficient of determination (R^2) were calculated, as shown in Table III. Checking the three indices, the FSTF predictions present greater accuracy generally. More precisely, the MAEs of FSTF are 0.683 K, 1.273 K, and 2.447 K smaller than those for STF for Regions 1 to 3, respectively. Furthermore, FSTF produces RMSEs that are 0.666, 0.900, and 2.189 K smaller than for STF for the three regions. Thus, when reconstructing LST time-series with large spatial gaps, FSTF can produce greater accuracy.

TABLE III
ACCURACY OF RECONSTRUCTION OF THE MODIS LST TIME-SERIES

	MAE (K)		RMSE (K)		R^2	
	STF	FSTF	STF	FSTF	STF	FSTF
Region 1	4.573	**3.890**	5.612	**4.946**	0.492	**0.658**
Region 2	4.689	**3.416**	5.429	**4.529**	0.800	**0.828**
Region 3	9.871	**7.424**	10.894	**8.705**	**0.777**	0.744

IV. DISCUSSION

A. *Prediction of 1 Km Hourly MODIS LST Data*

As presented in the Introduction, the CLDAS product can provide 7 km hourly LST. To reconstruct all-sky LST, this article utilized the CLDAS LST at the same acquisition time of MODIS LST for reconstruction. Ultimately, MODIS LST time-series were generated by FSTF. Although this research produces 1 km spatial resolution daily LST, the temporal resolution may be coarse for studies on diurnal variation in LST. Actually, with the reconstructed daily 1 km MODIS LST and 7 km hourly CLDAS LST, there exists a great possibility to obtain hourly MODIS-like LST by inheriting the spatial resolution of MODIS LST and temporal resolution of CLDAS. This process can be realized directly by employing spatio-temporal fusion. Once the spatially complete MODIS-CLDAS LST image pair at one time point in a day is acquired, it can be regarded as the known fine-coarse image pair. Thus, by fusing with CLDAS at other times during the day, 1 km LST for the other 23 h in a day can be predicted.

Taking the reconstruction of hourly 1 km LST on May 1 to May 10 for Region 1 as an example, the reconstruction results are shown in Fig. 11. To obtain a more reliable prediction, the former and latter temporally closest MODIS-CLDAS LST image pairs were applied for spatio-temporal fusion. For example, when reconstructing the 1 km LST at UTC 7:00 on May 2, the MODIS-CLDAS LST image pairs at UTC 4:00 on May 2 and May 3, together with the CLDAS at UTC 7:00 on May 2 were included. Checking the reconstruction results, the variation of LST within a day is reconstructed, as the LST increases from UTC 0:00 to 6:00 and decreases from UTC 6:00 to 23:00. Also, the variation of LST presents temporal continuity, which is in accordance with the common expectation that LST changes gradually over time. Generally, the reconstruction of 1 km hourly LST is feasible from visual inspection. The reconstructed 1 km hourly LST images

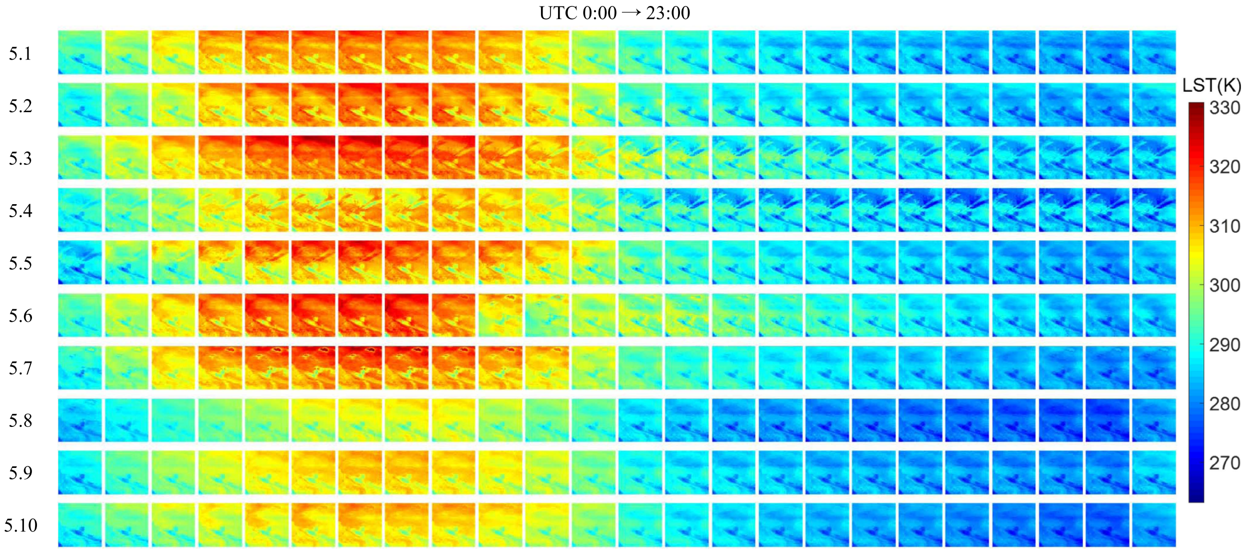


Fig. 11. Prediction of hourly MODIS LST data.

have great potential for research on diurnal variation in LST across different land cover types. Note that when generating 1 km hourly LST, the errors may accumulate and, thus, it is important to ensure a high accuracy of any previous daily LST reconstruction based on FSTF.

B. *Flexibility of FSTF*

In this article, a FSTF method is proposed to reconstruct all-sky MODIS LST images with the assistance of CLDAS data. FSTF includes two steps: gap filling and spatio-temporal fusion. By integrating a typical gap filling method (i.e., MNSPI) and one of the four spatio-temporal fusion methods (i.e., STARFM, ESTARFM, SPSTFM, and VIPSTF-SW), four specific forms of FSTF were developed in this article. As the two parts together determine the accuracy of FSTF, to further increase the reconstruction accuracy, it is of great need to explore more forms of FSTF by combining more powerful gap filling and spatio-temporal fusion methods. For gap filling, more spatial-temporal information-based methods can be considered, such as deep learning-based methods. The key issue would be to collect sufficient reliable training data based on the platform of a high performance computer. For spatio-temporal fusion, the four algorithms included in this article are spatial weighting-based methods and machine learning-based methods. In future research, the performance of FSTF may be improved by developing other spatio-temporal fusion methods, such as hybrid methods and deep learning-based methods. In potential models, it would be crucial to account for the change pattern of LST over time. Moreover, it is noted that the FSTF method proposed in this article implements gap filling and spatio-temporal fusion with images at just one or two time points. Actually, there exists great spatial and temporal correlation between the image time-series to be reconstructed. In future research, an extended FSTF version integrating the information of the image time-series deserves to be developed for more reliable reconstruction.

C. *Potential of FSTF*

The FSTF method proposed in this article aims at reconstructing large areas of data loss in MODIS LST, thus, contributing to the generation of all-sky MODIS LST products. It has great potential for updating the current MODIS LST product at the global scale. Furthermore, FSTF has the potential to be applied to more situations. First, FSTF can help to generate all-sky LST with finer spatial resolution by blending the predicted all-sky MODIS LST product with a finer spatial resolution, but coarser temporal resolution product. For example, the Landsat-8 TIR band can provide 100 m spatial resolution LST every 16 days, but also encounters the problem of spatial information loss. In this case, FSTF can be applied to reconstruct Landsat LST with data loss by fusing with all-sky 1 km MODIS LST, thus, generating all-sky 100 m LST. Second, other than LST, FSTF has the potential to support the reconstruction of other surface observation data. Generally, FSTF has the ability to reconstruct fine spatial resolution products with data loss by fusing with spatially complete products of the same type, but with coarser spatial resolution and finer temporal resolution. Actually, missing data is a common problem in surface observation products, as many products are generated from optical remote sensing images, which always face the issue of cloud contamination.

D. *Uncertainty in FSTF and Validation*

The FSTF method proposed in this article provides a new approach for reconstructing MODIS LST images. However, it is a method composed of multiple steps and, importantly, its performance relies heavily on the pregap filling process. As MNSPI cannot produce a perfect prediction, the error caused by gap filling may propagate to the postspatio-temporal fusion step. Thus, the impact of the error caused by the pregap filling process should be considered when applying spatio-temporal fusion. Moreover, it would also be worthwhile to explore a one-stage method that can realize gap filling and spatio-temporal fusion in a unified framework, where the uncertainty can be handled jointly.

In situ time-series data were employed for accuracy validation of FSTF in the temporal domain, due to a lack of real, spatially complete reference data. In practice, due to the lack of alternative data, in situ measurements serve as a common choice for evaluation of predicted LST time-series [58], [59]. However, there may exist errors in the reference data themselves, not least since the automatic meteorological stations are installed on towers instead of on the ground. Thus, when using in situ LST data to evaluate the accuracy of FSTF, uncertainties remain and these should be further investigated. In future research, ground measurements with greater reliability are expected to be explored to provide a more critical validation system.

FSTF provides a new means for making full utilization of the available data, and in this research was demonstrated to be more accurate than the conventional spatio-temporal fusion methods used widely for LST reconstruction in recent studies [8], [12], [19]. Moreover, the MODIS LST time-series reconstructed by FSTF tends to be closer to the in situ time-series data in the temporal trend [see Fig. 8(a), (c), and (e)]. Future research should explore ways to increase the accuracy of FSTF further to meet various research demands. In particular, in future research it would be interesting to further increase the accuracy of FSTF by involving multisource data and optimizing the integration of gap filling and spatio-temporal fusion methods.

V. Conclusion

As an important sensor for global monitoring, MODIS can provide 1 km LST every day, but is affected by different degrees of spatial information loss. For generation of an all-sky MODIS LST product, it is both necessary, and a great challenge, to reconstruct images with large gaps, especially for completely missing data in a local area of interest. This article proposes a FSTF method for reconstructing MODIS LST time-series with the assistance of CLDAS LST data. By integrating effectively gap filling and spatio-temporal fusion methods, FSTF provides a practical solution for all-sky MODIS LST time-series generation. Based on the experiments conducted in three regions, the following main conclusions are made.

1) FSTF is able to take full advantage of temporally close MODIS LST data with small gaps and can produce greater reconstruction accuracy than the original spatio-temporal fusion.
2) CLDAS is a type of effective auxiliary data for reconstructing MODIS LST with large gaps.
3) VIPSTF-SW-based FSTF is generally superior to STARFM-based and ESTARFM-based FSTF.

References

[1] A. Siddiqui, G. Kushwaha, B. Nikam, S. K. Srivastav, A. Shelar, and P. Kumar, "Analysing the day/night seasonal and annual changes and trends in land surface temperature and surface urban heat island intensity (SUHII) for Indian cities," *Sustain. Cities Soc.*, vol. 75, 2021, Art. no. 103374.

[2] X. Zheng, Z.-L. Li, T. Wang, H. Huang, and F. Nerry, "Determination of global land surface temperature using data from only five selected thermal infrared channels: Method extension and accuracy assessment," *Remote Sens. Environ.*, vol. 268, 2022, Art. no. 112774.

[3] B. Li et al., "Estimation of all-sky 1 km land surface temperature over the conterminous United States," *Remote Sens. Environ.*, vol. 266, 2021, Art. no. 112707.

[4] T. Hu et al., "Monitoring agricultural drought in Australia using MTSAT-2 land surface temperature retrievals," *Remote Sens. Environ.*, vol. 236, 2020, Art. no. 111419.

[5] F. Xie and H. Fan, "Deriving drought indices from MODIS vegetation indices (NDVI/EVI) and land surface temperature (LST): Is data reconstruction necessary?," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 101, 2021, Art. no. 102352.

[6] V. M. Bindhu, B. Narasimhan, and K. P. Sudheer, "Development and verification of a non-linear disaggregation method (NL-DisTrad) to downscale MODIS land surface temperature to the spatial scale of Landsat thermal data to estimate evapotranspiration," *Remote Sens. Environ.*, vol. 135, pp. 118–129, 2013.

[7] Y. Bai, N. Bhattarai, K. Mallick, S. Zhang, T. Hu, and J. Zhang, "Thermally derived evapotranspiration from the surface temperature initiated closure (STIC) model improves cropland GPP estimates under dry conditions," *Remote Sens. Environ.*, vol. 271, 2022, Art. no. 112901.

[8] H. Shen, L. Huang, L. Zhang, W. Penghai, and C. Zeng, "Long-term and fine-scale satellite monitoring of the urban heat island effect by the fusion of multi-temporal and multi-sensor remote sensed data: A 26-year case study of the city of Wuhan in China," *Remote Sens. Environ.*, vol. 172, pp. 109–125, 2016.

[9] Q. Meng, L. Zhang, Z. Sun, F. Meng, L. Wang, and Y. Sun, "Characterizing spatial and temporal trends of surface urban heat island effect in an urban main built-up area: A 12-year case study in Beijing, China," *Remote Sens. Environ.*, vol. 204, pp. 826–837, 2018.

[10] M. Wang et al., "A radiance-based split-window algorithm for land surface temperature retrieval: Theory and application to MODIS data," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 76, pp. 204–217, 2019.

[11] H. Wang et al., "A method for land surface temperature retrieval based on model-data-knowledge-driven and deep learning," *Remote Sens. Environ.*, vol. 265, 2021, Art. no. 112665.

[12] Q. Weng, P. Fu, and F. Gao, "Generating daily land surface temperature at Landsat resolution by fusing Landsat and MODIS data," *Remote Sens. Environ.*, vol. 145, pp. 55–67, 2014.

[13] H. Shafizadeh-Moghadam, Q. Weng, H. Liu, and R. Valavi, "Modeling the spatial variation of urban land surface temperature in relation to environmental and anthropogenic factors: A case study of Tehran, Iran," *GIScience Remote Sens.*, vol. 57, no. 4, pp. 483–496, 2020.

[14] M. D. King, S. Platnick, W. P. Menzel, S. A. Ackerman, and P. A. Hubanks, "Spatial and temporal distribution of clouds observed by MODIS onboard the Terra and Aqua satellites," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 3826–3852, Jul. 2013.

[15] H. Shen et al., "Missing information reconstruction of remote sensing data: A technical review," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 61–85, Sep. 2015.

[16] J. Gao, Q. Yuan, J. Li, H. Zhang, and X. Su, "Cloud removal with fusion of high resolution optical and SAR images using generative adversarial networks," *Remote Sens.*, vol. 12, no. 1, 2020, Art. no. 191.

[17] C. Shi, Z. Xie, H. Qian, M. Liang, and X. Yang, "China land soil moisture EnKF data assimilation based on satellite remote sensing data," *Sci. China Earth Sci.*, vol. 54, no. 9, pp. 1430–1440, 2011.

[18] M. Rodell et al., "The global land data assimilation system," *Bull. Amer. Meteorological Soc.*, vol. 85, no. 3, pp. 381–394, 2004.

[19] D. Long et al., "Generation of MODIS-like land surface temperatures under all-weather conditions based on a data fusion approach," *Remote Sens. Environ.*, vol. 246, 2020, Art. no. 111863.

[20] X. Zhu, F. Cai, J. Tian, and T. K.-A. Williams, "Spatiotemporal fusion of multisource remote sensing data: Literature survey, taxonomy, principles, applications, and future directions," *Remote Sens.*, vol. 10, no. 4, 2018, Art. no. 527.

[21] M. Belgiu and A. Stein, "Spatiotemporal image fusion in remote sensing," *Remote Sens.*, vol. 11, no. 7, 2019, Art. no. 818.

[22] B. Chen and B. Huang, "Comparison of spatiotemporal fusion models: A review," *Remote Sens.*, vol. 7, no. 2, pp. 1798–1835, 2015.

[23] F. Gao et al., "Fusing Landsat and MODIS data for vegetation monitoring," *IEEE Geosci. Remote Sens. Mag.*, vol. 3, no. 3, pp. 47–60, Sep. 2015.

[24] H. Zhang, B. Huang, M. Zhang, K. Cao, and L. Yu, "A generalization of spatial and temporal fusion methods for remotely sensed surface parameters," *Int. J. Remote Sens.*, vol. 36, pp. 4411–4445, 2015.

[25] J. Zhou et al., "Sensitivity of six typical spatiotemporal fusion methods to different influential factors: A comparative study for a normalized difference vegetation index time series reconstruction," *Remote Sens. Environ.*, vol. 252, 2021, Art. no. 112130.

[26] F. Gao, J. Masek, M. Schwaller, and F. Hall, "On the blending of the Landsat and MODIS surface reflectance: Predicting daily Landsat surface reflectance," *IEEE Trans. Geosci. Remote Sens.*, vol. 44, no. 8, pp. 2207–2218, Aug. 2006.

[27] X. Zhu, J. Chen, F. Gao, X. Chen, and J. G. Masek, "An enhanced spatial and temporal adaptive reflectance fusion model for complex heterogeneous regions," *Remote Sens. Environ.*, vol. 114, no. 11, pp. 2610–2623, 2010.

[28] Q. Wang and P. M. Atkinson, "Spatio-temporal fusion for daily Sentinel-2 images," *Remote Sens. Environ.*, vol. 204, pp. 31–42, 2018.

[29] T. Hilker et al., "A new data fusion model for high spatial- and temporal-resolution mapping of forest disturbance based on Landsat and MODIS," *Remote Sens. Environ.*, vol. 113, no. 8, pp. 1613–1627, 2009.

[30] Q. Wang, Y. Tang, X. Tong, and P. M. Atkinson, "Virtual image pair-based spatio-temporal fusion," *Remote Sens. Environ.*, vol. 249, 2020, Art. no. 112009.

[31] B. Zhukov, D. Oertel, F. Lanzl, and G. Reinhackel, "Unmixing-based multisensor multiresolution image fusion," *IEEE Trans. Geosci. Remote Sens.*, vol. 37, no. 3, pp. 1212–1226, May 1999.

[32] J. Amorós-López et al., "Multitemporal fusion of Landsat/TM and ENVISAT/MERIS for crop monitoring," *Int. J. Appl. Earth Observ. Geoinformation*, vol. 23, pp. 132–141, 2013.

[33] C. Gevaert and F. García-Haro, "A comparison of STARFM and an unmixing-based algorithm for Landsat and MODIS data fusion," *Remote Sens. Environ.*, vol. 156, pp. 34–44, 2015.

[34] R. Zurita-Milla, G. Kaiser, J. G. P. W. Clevers, W. Schneider, and M. E. Schaepman, "Downscaling time series of MERIS full resolution data to monitor vegetation seasonal dynamics," *Remote Sens. Environ.*, vol. 113, pp. 1874–1885, 2009.

[35] M. Wu, Z. Niu, C. Wang, C. Wu, and L. Wang, "Use of MODIS and Landsat time series data to generate high-resolution temporal synthetic Landsat data using a spatial and temporal reflectance fusion model," *J. Appl. Remote Sens.*, vol. 6, no. 1, 2012, Art. no. 063507.

[36] X. Zhu, E. H. Helmer, F. Gao, D. Liu, J. Chen, and M. A. Lefsky, "A flexible spatiotemporal method for fusing satellite images with different resolutions," *Remote Sens. Environ.*, vol. 172, pp. 165–177, 2016.

[37] M. Liu et al., "An improved flexible spatiotemporal DAta fusion (IFSDAF) method for producing high spatiotemporal resolution normalized difference vegetation index time series," *Remote Sens. Environ.*, vol. 227, pp. 74–89, 2019.

[38] X. Li et al., "SFSDAF: An enhanced FSDAF that incorporates sub-pixel class fraction change information for spatio-temporal image fusion," *Remote Sens. Environ.*, vol. 237, 2020, Art. no. 111537.

[39] D. Guo, W. Shi, M. Hao, and X. Zhu, "FSDAF 2.0: Improving the performance of retrieving land cover changes and preserving spatial details," *Remote Sens. Environ.*, vol. 248, 2020, Art. no. 111973.

[40] H. Song and B. Huang, "Spatiotemporal reflectance fusion via sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 10, pp. 3707–3716, Oct. 2012.

[41] V. Moosavi, A. Talebi, M. H. Mokhtari, S. R. F. Shamsi, and Y. Niazi, "A wavelet-artificial intelligence fusion approach (WAIFA) for blending Landsat and MODIS surface temperature," *Remote Sens. Environ.*, vol. 169, pp. 243–254, 2015.

[42] H. Song, Q. Liu, G. Wang, R. Hang, and B. Huang, "Spatiotemporal satellite image fusion using deep convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 3, pp. 821–829, Mar. 2018.

[43] Z. Tan, P. Yue, L. Di, and J. Tang, "Deriving high spatiotemporal remote sensing images using deep convolutional network," *Remote Sens.*, vol. 10, no. 7, 2018, Art. no. 1066.

[44] Z. Tan, L. Di, M. Zhang, L. Guo, and M. Gao, "An enhanced deep convolutional model for spatiotemporal image fusion," *Remote Sens.*, vol. 11, no. 24, 2019, Art. no. 2898.

[45] G. Yang et al., "MSFusion: Multistage for remote sensing image spatiotemporal fusion based on texture transformer and convolutional neural network," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 4653–4666, Jun. 2022.

[46] Z. Tan, M. Gao, X. Li, and L. Jiang, "A flexible reference-insensitive spatiotemporal fusion model for remote sensing images using conditional generative adversarial network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Jan. 2022, Art. no. 5601413.

[47] H. Zhang, Y. Song, C. Han, and L. Zhang, "Remote sensing image spatiotemporal fusion using a generative adversarial network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 5, pp. 4273–4286, May 2021.

[48] J. Chen, L. Wang, R. Feng, P. Liu, W. Han, and X. Chen, "CycleGAN-STF: Spatiotemporal fusion via CycleGAN-based image generation," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5851–5865, Jul. 2021.

[49] C. Zeng, H. Shen, and L. Zhang, "Recovering missing pixels for Landsat ETM + SLC-off imagery using multi-temporal regression analysis and a regularization method," *Remote Sens. Environ.*, vol. 131, pp. 182–194, 2013.

[50] Q. Cheng, H. Shen, L. Zhang, Q. Yuan, and C. Zeng, "Cloud removal for remotely sensed images by similar pixel replacement guided with a spatio-temporal MRF model," *ISPRS J. Photogrammetry Remote Sens.*, vol. 92, pp. 54–68, 2014.

[51] J. Chen, X. Zhu, J. E. Vogelmann, F. Gao, and S. Jin, "A simple and effective method for filling gaps in Landsat ETM+ SLC-off images," *Remote Sens. Environ.*, vol. 115, no. 4, pp. 1053–1064, 2011.

[52] X. Zhu, F. Gao, D. Liu, and J. Chen, "A modified neighborhood similar pixel interpolator approach for removing thick clouds in Landsat images," *IEEE Geosci. Remote Sens. Lett.*, vol. 9, no. 3, pp. 521–525, May 2012.

[53] Q. Zhang, Q. Yuan, Z. Li, F. Sun, and L. Zhang, "Combined deep prior with low-rank tensor SVD for thick cloud removal in multitemporal images," *ISPRS J. Photogrammetry Remote Sens.*, vol. 177, pp. 161–173, 2021.

[54] Q. Zhang, Q. Yuan, J. Li, Z. Li, H. Shen, and L. Zhang, "Thick cloud and cloud shadow removal in multitemporal imagery using progressively spatio-temporal patch group deep learning," *ISPRS J. Photogrammetry Remote Sens.*, vol. 162, pp. 148–160, 2020.

[55] Q. Wang, L. Wang, X. Zhu, Y. Ge, X. Tong, and P. M. Atkinson, "Remote sensing image gap filling based on spatial-spectral random forests," *Sci. Remote Sens.*, vol. 5, 2022, Art. no. 100048.

[56] S. M. Liu et al., "A comparison of eddy-covariance and large aperture scintillometer measurements with respect to the energy balance closure problem," *Hydrol. Earth Syst. Sci.*, vol. 15, no. 4, pp. 1291–1306, 2011.

[57] S. M. Liu, Z. W. Xu, Z. L. Zhu, Z. Z. Jia, and M. J. Zhu, "Measurements of evapotranspiration from eddy-covariance systems and large aperture scintillometers in the Hai River Basin, China," *J. Hydrol.*, vol. 487, pp. 24–38, 2013.

[58] W. Wang, S. Liang, and T. Meyers, "Validating MODIS land surface temperature products using long-term nighttime ground measurements," *Remote Sens. Environ.*, vol. 112, pp. 623–635, 2008.

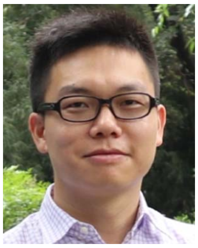
[59] P. Wu et al., "Spatially continuous and high-resolution land surface temperature product generation: A review of reconstruction and spatiotemporal fusion techniques," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 3, pp. 112–137, Sep. 2021.

[60] K. Wang, P. Wang, M. Sparrow, J. Liu, X. Zhou, and S. Haginoya, "Estimation of surface long wave radiation and broadband emissivity using moderate resolution imaging spectroradiometer (MODIS) land surface temperature/emissivity products," *J. Geophysical Res.*, vol. 110, 2005, Art. no. 11109.

[61] K. Wang and S. Liang, "Evaluation of ASTER and MODIS land surface temperature and emissivity products using long-term surface longwave radiation observations at SURFRAD sites," *Remote Sens. Environ.*, vol. 113, no. 7, pp. 1556–1565, 2009.

Yijie Tang received the B.S. degree from Nanjing Normal University, Nanjing, China, in 2019. She is currently working toward the Ph.D. degree with Tongji University, Shanghai, China.

Her research interests focus on remote sensing image fusion.



Qunming Wang received the Ph.D. degree from The Hong Kong Polytechnic University, Hong Kong, in 2015.

He is currently a Professor with the College of Surveying and Geo-Informatics, Tongji University, Shanghai, China. From 2017 to 2018, he was a Lecturer (Assistant Professor) with Lancaster Environment Centre, Lancaster University, Lancaster, U.K., where he is currently a Visiting Professor. His 3-year Ph.D. study was supported by the hypercompetitive Hong Kong Ph.D. Fellowship and his Ph.D. thesis was awarded as the Outstanding Thesis in the Faculty. He has authored or coauthored more than 70 peer-reviewed articles in international journals such as *Remote Sensing of Environment*, IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, and *ISPRS Journal of Photogrammetry and Remote Sensing*. His research interests include remote sensing, image processing, and geostatistics.

Prof. Wang is an Editorial Board Member for *Remote Sensing of Environment*, and serves as an Associate Editor for *Science of Remote Sensing* (sister journal of *Remote Sensing of Environment*) and *Photogrammetric Engineering & Remote Sensing*. He was an Associate Editor for *Computers and Geosciences* (2017−2020).

Peter M. Atkinson received the Ph.D. degree from The University of Sheffield, Sheffield, U.K. (NERC CASE award with Rothamsted Experimental Station), in 1990, and the MBA degree from the University of Southampton, Southampton, U.K., in 2012.

He is currently a Distinguished Professor of Spatial Data Science and Dean of the Faculty of Science and Technology, Lancaster University, Lancaster, U.K. He was previously a Professor of Geography with the University of Southampton, where he is currently a Visiting Professor. He is also a Visiting Professor with the Chinese Academy of Sciences, Beijing, China. He previously held the Belle van Zuylen Chair with Utrecht University, The Netherlands. He has authored or coauthored more than 300 peer-reviewed articles in international scientific journals and around 50 refereed book chapters. He has also edited nine journal special issues and eight books. The main focus of his research is in remote sensing, geographical information science and spatial (and space-time) statistics applied to a range of environmental science and socio-economic problems.

Prof. Atkinson is the recipient of the Peter Burrough Award of the International Spatial Accuracy Research Association and is a Fellow of the Learned Society of Wales. He is the Editor-in-Chief for *Science of Remote Sensing*, a sister journal of Remote Sensing of Environment. He also sits on the editorial boards of several further journals including Geographical Analysis, Spatial Statistics, International Journal of Applied Earth Observation and Geoinformation, and Environmental Informatics.